

# PLAUSIBLE AUDITORY AUGMENTATION OF PHYSICAL INTERACTION

Marian Weger<sup>1</sup>, Thomas Hermann<sup>2</sup>, Robert Höldrich<sup>1</sup>

<sup>1</sup> IEM, University of Music and Performing Arts, Graz, Austria

<sup>2</sup> Ambient Intelligence Group, CITEC, Bielefeld University, Bielefeld, Germany  
weger@iem.at

## ABSTRACT

Interactions with physical objects usually evoke sounds, i.e., auditory feedback that depends on the interacting objects (e.g., table, hand, or pencil) and interaction type (e.g., tapping or scratching). The continuous real-time adaptation of sound during interaction enables the manipulation/refinement of perceived characteristics (size, material) of physical objects. Furthermore, when controlled by unrelated external data, the resulting ambient sonifications can keep users aware of changing data. This article introduces the concept of *plausibility* to the topic of auditory augmentations of physical interactions, aiming at providing an experimentation platform for investigating surface-based physical interactions, understanding relevant acoustic cues, redefining these via auditory augmentation / blended sonification and particularly to empirically measure the plausibility limits of such auditory augmentations. Besides conceptual contributions along the trade-off between plausibility and usability, a practical experimentation system is introduced, together with a very first qualitative pilot study.

## 1. INTRODUCTION

As we live in a physical world, we interact with other physical objects: the floor we are walking on, objects we take in our hands, or a table where such objects may be positioned. Any of such physical interactions usually evoke specific sounds, depending on the physical properties of the involved objects, such as material, shape, or size, and also depending on the type of interaction, i.e., excitation (scratching, hitting, etc.). Physical objects in this context refer to objects with spatial location [1]. The resulting sounds being a reaction to a preceding (physical) action are referred to as auditory feedback, representing a major part of our everyday acoustic environment which we use either consciously or unconsciously while pursuing our daily activities.

According to activity theory, these activities can be separated into a number of goal-oriented conscious actions, which in turn consist of chains of unconscious operations [2, 3]. One possible action using auditory feedback may be to shake a box in order to get information on its contents from the resulting sound. Similarly, our operations are influenced by auditory feedback, for example, through footstep sounds which reveal important information on the material of the ground and thus subconsciously influence our walking style [4, 5, 6, 7].

In this context, we define *augmented auditory feedback* as the artificially modified sonic reaction to physical interaction. This type of augmentation could be exploited in various forms: First, the signal-to-noise-ratio may be improved in order to help communication of (a) the involved objects' physical properties such as material or spatial dimensions, and (b) the interaction type such as tapping or scratching. Both may facilitate specific activities. Second, specific physical properties of a physical object could be modified perceptually, in order to induce a certain change in user behavior. Apart from the aforementioned footstep sounds, successful behavioral change through sound has been shown for hand tapping [8, 9], as well as by for grasping [10, 11]. Finally, augmented auditory feedback creates a new communication channel that can also be used for sonification of data that are completely unrelated to both object and interaction, e.g., for continuous monitoring as a secondary task. The sonification is thus naturally and seamlessly fitted into the everyday acoustic environment.

In the context of this article, we explore the exploitability of augmented auditory feedback as communication channel. In information theory, a major characteristic of a given channel is its channel capacity, i.e., the highest rate at which information can be reliably transmitted. In our case, it is limited by at least two factors: the *plausibility* and the *usability* of augmented auditory feedback. Both will be discussed in detail in Sec. 2.

For a specific physical interaction, we make four assumptions:

- (1) There exists a manifold of sounds which serve as *plausible auditory feedback*. Its borders define the *plausibility range*.
- (2) There exists a manifold of sounds which serve as *usable auditory feedback*, i.e., the sounds are helpful to perform the specific actions. Its borders define the *usability range*.
- (3) The manifolds of plausible and usable sounds overlap. We define this overlap region as the manifold of *alternative auditory feedbacks*.
- (4) It is possible to discriminate between different alternative auditory feedbacks.

If all these assumptions are met (compare Fig. 1), we conclude that it is possible to convey additional information through (more or less subtle) sound changes within the intersection of plausible and usable auditory feedbacks.

Note that usability refers to the original physical interaction alone. Adding additional information can make an interaction more useful for the users and eventually let users even actively interact to query information that they regard as useful. This aspect of usability (or the additionally received information) is not meant here.

The above argumentation leads us directly to the already established concept of *auditory augmentation* [12]. In addition, we



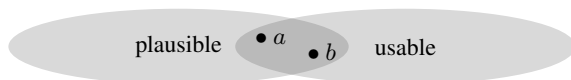


Figure 1: Sets of plausible and usable variants of auditory feedback for a specific physical interaction. Points *a* and *b* represent two discriminable but still plausible and usable sounds.

contribute a more systematic design approach to the no less important original definition. According to [12], auditory augmentation is “a paradigm to vary the objects’ sonic characteristics such that their original sonic response appears as augmented by an artificial sound that encodes information about external data. [...] All this manipulation does not affect the sound’s original purpose.”

In this article, we test the method of auditory augmentation through the concept of plausible and usable auditory feedback and provide an exemplary implementation as proof of concept. What follows in Sec. 2 is an investigation of the above assumptions concerning plausible and usable auditory feedback. After a review of related sonification literature in Sec. 3, we present an experimental platform for an auditory augmentation system in Sec. 4, together with a pilot study with naive users.

## 2. PLAUSIBLE AND USABLE AUGMENTED AUDITORY FEEDBACK

In this section, we test the four assumptions formulated in the introduction in order to find rules how to meet them.

### 2.1. Plausible auditory feedback

*“I’m not concerned with plausibility;  
that’s the easiest part of it, so why bother?”*

— Alfred Hitchcock [13, p. 99]

Sensory feedback is generally considered to be plausible if it is “conceptually consistent with what is known to have occurred in the past” [14]. In particular, “a highly plausible scenario is one that fits prior knowledge well: with many different sources of corroboration, without complexity of explanation, and with minimal conjecture” [14].

With increasing perceptual dominance of auditory augmentation, the resulting auditory feedback influences user perception, emotion, and behavior [9]. For example, auditory cues have been shown to influence the haptic perception of virtual textures presented by the Phantom Omni haptic device [15]. Likewise, perception of material properties (e.g., hard/soft, rough/smooth) has been shown to be strongly influenced by auditory cues [16]. As the perceptual plausibility depends on the congruency between different modalities such as haptic, visual, or auditory information, it has no meaning for the unimodal case of auditory feedback alone. Perceptual congruency and therefore plausibility is high if the information of the different modalities combined, i.e., the combination of different stimuli, matches the pattern we learned through natural interactions with our physical environment. It is therefore hypothesized that perceptual plausibility increases with increasing congruency (agreement) between cues (information) from different sensory modalities (information channels).

It must be considered that people are already accustomed to manipulated visual feedback, but generally have less experience (or none at all) in augmented auditory or tactile feedback: It is

common knowledge that a physical object’s inner structure can be concealed, e.g., through painted surfaces. This supports the assumption that there exist at least several interchangeably plausible visual representations of a physical object.

We argue that if an auditory augmentation alters the perception of only such physical properties that are hidden behind the surface finishing (e.g., lacquer or laminate), i.e., physical properties that are not conveyed through vision or haptics, then the auditory augmentation cannot lead to incongruent sensory information. The augmented physical properties, however, still need to remain physically viable in order to reduce the complexity of explanation. These physical properties concerning the inner structure of physical objects may include material category, density, hollowness, and spatial volume, as well as boundary condition (e.g., free or clamped) and coupling to other physical objects.

Individual exploitability of the mentioned physical properties for sonification purposes can be derived from a short literature review on auditory perception:

*Material identification.* Listening experiments showed that material identification of impacted plates is limited to gross density categories (steel/glass vs. wood/plexiglass), while in-category identification is strongly influenced by interpretations based on the everyday acoustic environment [17]. In particular, glass and wood were associated with smaller objects than metal and plastic.

*Volume of cavities.* It has been shown that volumes of cavities can be directly perceived through sound [18].

*Hollowness.* Participants of a listening experiment could auditorily distinguish between hollow and solid bars; however, they tended to base their judgments on simple differences in frequency, which accompany size variations of a resonant cavity [19].

*Size.* As a general rule when judging the size of a sound-producing object, people tend to rely on simple sound attributes such as loudness, duration, and base frequency [20, 21, 22].

*Boundary condition.* Concerning stiffness variations at different boundary conditions of vibrating plates, it has been shown that listeners used both pitch and loudness for discrimination [23].

Concerning other physical properties, the interested reader is referred to further literature on auditory perception of physical objects, as reviewed in [24, pp. 1–16] or [25, pp. 31–38].

Summing up, all of the above material properties have a perceivable effect on auditory feedback. In consequence, their modulation through auditory augmentation is supposed to provide a reliable communication channel. The resulting augmented auditory feedback is assumed to stay within the plausibility range if augmented in a physically meaningful and feasible way, i.e., if the illusory physical properties can be explained without effort.

Note that augmented auditory feedback that is far from physical feasibility, but rather represents a cartoonification of the original auditory feedback, might be perceived as even more plausible or realistic than the original sound [26]. However, as we seek for general applications, the more restricted approach of physically informed auditory augmentation is followed here.

### 2.2. Usable auditory feedback

According to ISO 9241-210 [27], usability is defined as the “extent to which a system, product or service can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use”.

Auditory feedback generally has a positive effect on task performance as well as on motor learning [28]. Sigrist et al. summa-

alized the potential and challenges of augmented feedback in general in the context of motor learning, and provided design criteria to successful visual, auditory, haptic, and multisensory feedback [28]. However, a positive effect of auditory feedback observed in isolation may completely vanish in the presence of feedback in other sensory modalities. For example, auditory feedback has been shown to improve the typing performance on a computer keyboard, if no additional haptic feedback was present [29], concluding that in this case, haptic feedback had stronger influence than auditory feedback.

In any case, what makes auditory feedback usable is the information it carries about the performed action and the physical objects it affects. All information which is encoded in the auditory feedback but not needed for a specific task, physical object, and context, is therefore assumed to be irrelevant for usability. We define *relevant physical properties* as the properties of a physical object which influence its usability. Likewise, *irrelevant physical properties* are the properties of a physical object which do not affect its usability.

Irrelevant physical properties are therefore considered as possible candidates for usability-independent auditory augmentation. Note that auditory augmentation of irrelevant physical properties may mask the perception of relevant physical properties – either auditorily, or through informational masking effects, e.g., by adding additional disturbance or stress. Furthermore, the relevance of specific physical properties of one physical object diverge for different actions and context.

As an example, we examine a relatively simple category of physical objects: an even, horizontal, rigid, and stationary surface. Such a surface usually appears on tables, cupboards, bookshelves, etc. Due to its affordances [30, pp. 10–13], it is primarily used for putting things on top, moving these things around, and manipulating them (writing, cooking, etc.), but also manual interaction is possible (hitting, scratching, tapping, etc.).

For the observed surface, we consider three relevant physical properties: Hardness influences how we put fragile things on top, or how we interact with our hands. Roughness influences how objects or fingers can be moved. Sturdiness is relevant, as a fragile table might break while positioning heavy things. We thereby do not want to modulate these but rather leave them unchanged in order to preserve usability.

Similarly, physical properties that we consider as irrelevant include spatial volume, hollowness, underlying material category, and boundary conditions – if not in conflict with relevant physical properties. These relate mainly to the non-visible part of the surface, under the visible texture layer. In consequence, their perception may be securely altered through auditory augmentation within the plausibility range.

Note that not all physical objects incorporate a specified purpose or intended use. However, every physical object has affordances. The goal of usable auditory augmentation is to preserve these in the best possible way while adding new affordances such as exploratory data analysis through manual interaction.

### 3. RELATED SONIFICATION WORK

With Wetterreim [12], Bovermann et al. already presented a proof-of-concept for the use of auditory augmentation in a peripheral monitoring situation where realtime weather data modulated the auditory feedback of a computer keyboard. Their general implementation processes an audio signal from a contact microphone

attached to a physical object by a digital filter whose parameters are controlled by external data. The modified signal is rendered by loudspeakers and blends with the object's original sound, providing an additional layer between a user's action and a physical object's auditory feedback.

The concept of auditory augmentation is strongly connected to *blended sonifications* which “blend into the users' environment without confronting users with any explicitly perceived technology” and thus provide ambient sonification channels [31]. This general framework contains auditory augmentation as an important special case (where user-caused audio signals affect the auditory display via data-dependent filtering). Yet more generally, in Blended Sonification, physical environment, digital environment, and user can each provide both audio and data components which affect the auditory display via filtering and/or additive auditory elements, bound together by the shared claim that the formed auditory gestalt is still perceived as coherent auditory events.

Another interesting yet more remotely relevant approach for the topic of this article was presented by Barrass [32], who embedded interactive sonifications into found objects, making these serve as a physical medium for exploration of the otherwise imperceptible and abstract data.

Interactive control of in-home ambient sonification systems through small and unobtrusive sensors has been discussed by [33]. The presented prototypes were able to convey information through sound, triggered by simple everyday actions.

Specifically connected to the affordances of surfaces is the sonification of pen strokes presented in [34]. Müller et al. used a simple physical model of a cylindrical resonator (pen) excited by a mainly stochastic process which describes the surface texture. They discriminate between three auditory surface textures: The micro surface texture consists of fine-grained single peaks of stochastic distribution which cannot be identified individually. The meso surface texture is rougher, resulting in isolated perceptible peaks. The macro surface texture relates to visible patterns, similar to a chessboard, and describes larger-scale and regular structures. The authors argue that “a maximum of naturalism of the sound feedback can be mainly achieved by providing a feedback to the interaction with low latency and high precision.” [34, p. 3].

### 4. THE MONDRIAN-TABLE

In order to explore the potential as well as the constraints of plausible and usable auditory augmentation, we developed an experimentation platform which adheres to the horizontal, even, rigid, and stationary surface, as exemplified in Sec. 2.2. For even further simplification, we restrict it to rectangular shape. The experimentation platform is intended to offer two different modes of operation. On the one hand, it should enable calm and unobtrusive blended sonification outside the focus of attention while performing daily activities that affect it (e.g., writing or positioning of other physical objects). On the other hand, it should serve as an interface for exploratory data analysis through manual interaction in the form of tapping, scratching, etc.

At the same time, irrelevant of the performed action, this experimentation platform is designed to offer plausible and usable augmented auditory feedback. As derived in Sec. 2.2, possible usability-independent physical properties include spatial volume (thickness of the plate), material category (wood, metal, etc.), and boundary conditions (free, clamped, simply supported, etc.). All these can be referred to the inner structure of the physical object

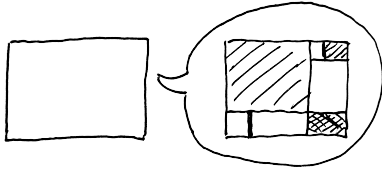


Figure 2: Concept of the Mondrian Table, inspired by Piet Mondrian’s “Composition II in Red, Blue, and Yellow” (1930).

and thus may be augmented in a plausible way, as described in Sec. 2.1.

Even on an ordinary dining table, the tabletop can be composed of several elements of different materiality. Such structures can be simulated through a space-dependent auditory augmentation. Our experimentation platform might therefore be perfectly suitable for sonification of a time-invariant configuration of geometric structures, displayed as regions with different perceived physical properties. This relates to a macro surface texture [34]. The underlying data can be a map, a technical drawing, or a painting similar to Piet Mondrian’s “Composition” series. We therefore call this experimentation platform the Mondrian Table (see Fig.2).

The Mondrian Table ambiently sonifies geometric structures as plausible and usable changes of the surface material. By placing a sheet of paper on top of it, those structures can be traced back with a pencil and thus transferred into the visual domain.

In contrast to the original approach of auditory augmentation, we insert a physical sound synthesis model between data and sonification, in the hope to achieve naturally plausible auditory feedback. The data does not directly produce sound, but is mapped to properties of the intermediate physical model (e.g., material category or shape). This model serves as a filter, altering the original sounds emerging from the interaction between user and physical object. Note that this is *not* Model-Based Sonification (MBS) [35, p. 399–427], but merely using physical models for perceptually more plausible intertwined coherences. However, the rendering pipeline makes our prototype as well perfectly suited for future interactions with MBS.

Tünnermann et al. [31] developed a standard for visualizing the audio- and dataflow of blended sonifications in a simple way. Such a blended sonification diagram usually involves the three factors User, Physical Environment, and Digital Environment as main sources of Data (D) and/or Audio (A). Connections of these sources to the Auditory Display as a sink visualize causal contributions to either a Filtered (F) or an Added (A) output. While filtered sonifications are “sonifications that stay very close to the original sound”, superimposed “sound samples or synthesized sound fall in the ‘added’ category” [31, p. 4].

The blended sonification diagram of the Mondrian Table (Fig.3) shows that our system produces a filtered sonification which is dependent on audio from the user in order to sonify data from the digital environment. The auditory display additionally depends on properties of the interaction (spatial location) as well as on properties of the physical object (original physical properties). We interpret the physical environment as something that is out of reach and not influenced by the user, and thus attribute properties of both interaction and involved physical object as data of the user.

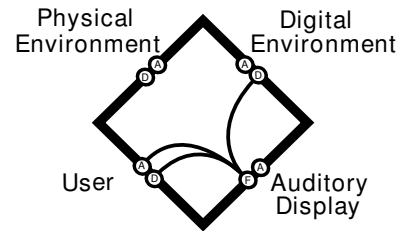


Figure 3: Blended sonification diagram of the Mondrian Table.

#### 4.1. Development of an experimentation platform

The technical implementation of our experimentation platform is similar to the augmented keyboard described in [12], replacing the keyboard by a rectangular plate with location tracking system. For the first prototype (see Fig. 4), we decided for a stylus-based graphic tablet which is basically a combination of both plate and tracking system. An advantage of the graphic tablet (compared to resistive sensitive surfaces) is, that the pen coordinates are already tracked while the pencil hovers contact-free over the surface, so that for instance filter parameters can be adjusted before any physical interaction happens. This is less critical for continuous interactions such as painting or scratching, but more so for tapping where the filter should have been parameterized *before* the pencil-paper interaction delivers the source signal for filtering. We use a Wacom Intuos 5 touch M which offers an active area of 224 by 140 mm – approximately the size of US Half Letter or A5 format, just enough for drawing simple sketches or writing small texts.

As we know that spatial congruency between sensory modalities has an impact on plausibility [36], sound playback needs to be somehow spatialized in a meaningful way. Horizontally, a landscape oriented sheet of size A5 within reach for writing comfortably covers an angle of approximately  $21^\circ$ . This is well below the usual  $60^\circ$  for simple stereophonic panning strategies with two loudspeakers. Considering a perceptual localization blur of  $\pm 3.6^\circ$  for white noise-bursts in the horizontal plane [37, p. 41], a stereophonic setup is supposed to be sufficient for the given task. Due to the small size of the sonification interface, vertical spatialization is regarded as irrelevant, taking into account the stronger perceptual localization blur in vertical angle ( $\pm 4^\circ$  for white noise, [37, p. 44]) as well as in distance perception ( $\pm 25\%$  at 1.1 m for impulses, [37, p. 47]). Contact microphones and loudspeakers are therefore placed vertically centered on the left and right side of the graphic tablet to provide natural panning in the case that both signal paths (left and right) are processed individually.

The latency of augmented auditory feedback needs to lie below a certain threshold in order to be successfully and plausibly combined with visual or haptic information, as also emphasized by Müller et al. [34] who referred to naturalism in this context. Typical maximum thresholds for synchronous perception are at 25 ms for auditory-tactile delay [38] and 18 to 24 ms [39, 40] for auditory-visual delay. However, for trained users such as musicians, a latency below 10 ms is recommended [38].

As hardware platform, we therefore decided for the Beagle-Bone Black Rev. C with Bela audio cape, which is designed for sub-millisecond-latency audio and sensor processing [41] and targeting specifically for digital musical instruments [42]. Two low-cost piezo-electric contact microphones are installed on the left and right side underneath the graphic tablet’s active area. These

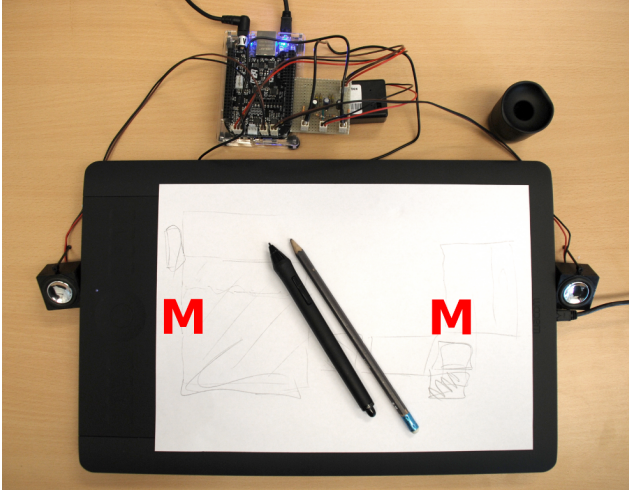


Figure 4: The Mondrian-Graphic-Tablet. The positions of the contact microphones underneath are marked through a red ‘M’.

are connected to the Bela cape through a FET buffer preamp (see [43] for schematics), driven by a 9 V battery. Two miniature loudspeakers are connected directly to the on-board class-D amplifiers of the Bela, and are placed besides the tablet.

For sound synthesis, in contrast to the sonification of pen strokes described in [34], we directly use the input signal from contact microphones as excitation signal for filter-based modal synthesis as described, e.g., in [44, pp. 46–48]. The physical model is intended to synthesize the normal modes (transversal waves) of a two-dimensional plate and to ‘apply’ these to the physical surface. The plate is seen as a linear and time-invariant system; its impulse response can be decomposed into exponentially decaying pure sine-waves [45, p. 12]:

$$x(t) = \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} A e^{-\alpha t} \sin(2\pi f_d t + \phi), \quad (1)$$

with mode indices  $m$  and  $n$  referring to the number of nodal lines in the two dimensions, respectively. Each mode is described by four factors: a frequency  $f_d$ , an attenuation rate  $\alpha$  for the exponential decay, an amplitude  $A$ , and a starting phase angle  $\phi$ , all dependent on mode indices  $m$  and  $n$ . In our simplified model, we assume amplitude and phase primarily influenced by the excitation signal and thus decided to ignore these in the first prototype.

As an example, we analyze an undamped, isotropic, and rectangular thin plate which is simply supported (hinged) along the edges. Its natural frequencies  $f_0$  are computed as follows [45, p. 80]:

$$f_0 = 0.453 h \sqrt{\frac{E}{\rho(1-\nu^2)}} \left[ \left( \frac{m+1}{l_x} \right)^2 + \left( \frac{n+1}{l_y} \right)^2 \right]. \quad (2)$$

In Eq. 2, the geometry of the plate is described by width  $l_x$ , height  $l_y$ , and thickness  $h$ . The material of the plate is expressed by Young’s modulus  $E$ , Poisson ratio  $\nu$ , and density  $\rho$ .

When damping is considered, the actual resonant frequencies  $f_d$ , i.e., the frequencies of the exponentially decaying sinusoids in Eq. 1, are a bit lower than the undamped natural frequencies  $f_0$ ,

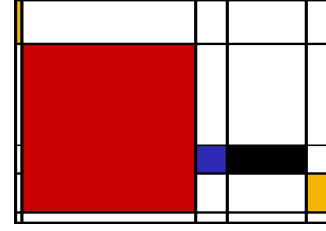


Figure 5: Example image, generated with the Mondrian Generator [48], serving as test data.

depending on the individual damping ratio  $\xi$  [45, p. 11]:

$$f_d = f_0 \sqrt{1 - \xi^2}, \quad \text{with} \quad \xi = \frac{\alpha}{2\pi f_0}. \quad (3)$$

The damping of a plate depends on several factors including geometry, material, and boundary condition, and cannot be easily predicted. We therefore employ a simple but perceptually-tuned damping model that is parametrized by global damping  $\alpha_G$  and frequency-relative damping  $\alpha_R$  [46]:

$$\alpha = e^{\alpha_G + \alpha_R \cdot 2\pi f_0}. \quad (4)$$

The resonant frequencies further depend on the boundary conditions of the plate edges. We describe the four boundary conditions through  $B_{x0}^i$  and  $B_{x1}^i$  for the left and right edge, and  $B_{y0}^i$  and  $B_{y1}^i$  for the top and bottom edge, respectively, where 0 means free, 1 means simply supported, and 2 means clamped. Solutions for all combinations of these are derived in [47].

In our simplified physical model, we neglect the resonant behavior of the graphic tablet and thus interpret the input signal from a contact microphone as a pure excitation signal induced by the stylus or pencil. This excitation signal is then filtered through our physical model to form the augmented output signal. Technically, this model is represented through a parallel filter-bank of resonant bandpass-filters which are tuned to the frequencies and attenuation rates of the simulated plate’s normal modes. The implementation is done in SuperCollider (SC), based on the DynKlank object which adds any number of resonances to an input signal, taking vectors of undamped natural frequencies  $f_0$ , amplitudes  $A$  (set constant to 1), and 60 dB decay times  $t_{60} = \ln(1000)\tau$  (in seconds, with time-constant  $\tau = 1/\alpha$ ) as arguments. The input signal is additionally cleaned from unwanted low-frequency noise through a high-pass filter set below the lowest mode frequency; the output signal is soft-limited.

Only the audio signal path is processed on the Bela running an SC server, while the mapping between data, physical model, and filter parameters is done on a separate computer running the SC language. Both computer and Bela communicate via network. The graphic tablet, connected via USB to the computer, is accessed as HID device.

In order to simplify the process of defining regions of different physical properties on the tablet already during development, a so-called Mondrian Generator [48] was used. This simple program generates compositions in red, blue, and yellow, ‘‘in the style of Piet Mondrian’’. The result is a raster graphics; an example is shown in Fig. 5.

Each of the three colors red, blue, and yellow represents a reference model  $M^i$  with specific physical properties, with  $i$  being

Table 1: Physical properties of the reference models.

	$M^1$	$M^2$	$M^3$
material category	metal	ceramics	glass
$E$ (GPa)	180	360	70
$\rho$ ( $kg\ m^{-3}$ )	7740	3800	2600
$\nu$	0.305	0.22	0.22
$\alpha_G$	0.6	1.55	2.0
$\alpha_R$ ( $\times 10^{-4}$ )	2	1.75	1.5

the index of the model. Each model  $M^i$  is represented by its parameter vector:

$$\mathbf{p}^i = [B_{x0}^i, B_{x1}^i, B_{y0}^i, B_{y1}^i, E^i, \rho^i, \nu^i, \alpha_G^i, \alpha_R^i, l_x^i, l_y^i, h^i].$$

The three reference models are rendered in parallel with pre-set parameter vectors  $\mathbf{p}_1$ ,  $\mathbf{p}_2$ , and  $\mathbf{p}_3$ , respectively. Together they form the set of possible auditory augmentations. The color directly controls the input gains  $g_{in}^1$ ,  $g_{in}^2$ , and  $g_{in}^3$  of the three models, respectively (red:  $g_{in}^1=1$ ,  $g_{in}^2=g_{in}^3=0$ ; blue:  $g_{in}^2=1$ ,  $g_{in}^1=g_{in}^3=0$ , yellow:  $g_{in}^3=1$ ,  $g_{in}^1=g_{in}^2=0$ ). White leads to a level-preserving mix of the three models. Black means that all input gains are set zero; black lines between the different regions are thus interpreted as grooves filled with sound-absorbing material.

#### 4.2. Informal evaluation

For evaluation of the prototype system, we defined three reference models with plate dimensions  $l_x$  and  $l_y$  set to match the active area of the graphic tablet. All three models were set to a constant thickness  $h$  of 3 mm, with the boundary conditions for all edges set free, i.e., simulating plates that were freely hovering in the air. The material properties, as listed in Tab. 1, were set to match the material categories metal, ceramics, and glass, respectively.

Due to limited computing power on the BeagleBone, only few resonances can be processed. However, the number of vibrational modes in the relevant audible range, i.e., below 16 kHz, is anyway small for the synthesized plates, especially due to the small spatial dimensions. In particular, these were 13 modes for metal (starting at 1.5 kHz), 6 modes for ceramics (starting at 3.0 kHz), and 13 modes for glass (starting at 1.6 kHz). For each model, only the 6 lowest out of 100 calculated modes have been used.

For informal evaluation, an unknown random image was created by the Mondrian Generator and sonified through the Mondrian Table by using the pre-set reference models. The graphic tablet was covered by a blank sheet of paper. The given task was to trace back the sonified structures with the stylus, and to draw a visual representation of the auditorily perceived structure. For this task, the stylus was paired with a real pen in order to be able to paint contours on the paper while exploring borders of a region of equal material properties. Such a sketch can be seen in Fig. 4.

An example video, recorded from first-person perspective with Soundman OKM binaural microphones, can be found in [49] (listening with headphones is recommended). It illustrates some tapping and scratching interactions with the pen at various locations of different data-driven (resp. location-dependent) augmentations.

The evaluation of the experimental platform provided us with valuable information on further development. In accordance with related studies, latency seems to be a major factor for plausible and successful auditory augmentation. During the evaluation, a round-trip latency of 25 ms was disturbing, while 14 ms was sufficient for the illusion of realistic auditory feedback when watching another

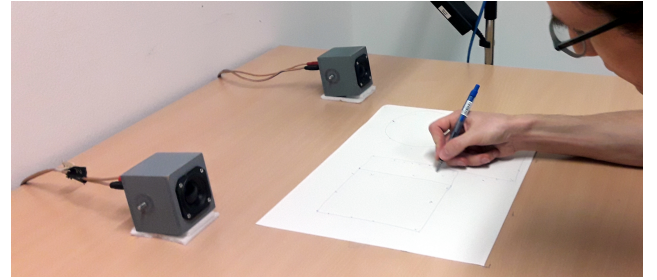


Figure 6: The Auditory Coloring Book.

person interacting (auditory-visual condition), but still felt unnatural for the interacting person (auditory-visual-haptic condition). With measured round-trip latency of about 1.5 ms, obtained with the Bela, the augmentation felt completely plausible and blended well with the direct sound of the original auditory feedback, despite spatially incongruent playback through loudspeakers. The synthesized materials felt realistic and successfully created the impression of a different materiality of the augmented physical surface. The sonified geometric shapes could be easily detected without effort while interacting in a completely natural way.

#### 4.3. The Auditory Coloring Book

In order to get an impression on how unbiased and naive users might interact with such an augmented surface, we installed a modified version of the above-described system at an open house event of the university: The Auditory Coloring Book (see Fig. 6) is an interactive sound installation which auditorily displays regions of different color through augmented surface material of a real table. In contrast to the first prototype, this installation uses a real table equipped with two AKG C411 contact microphones underneath the surface. Two miniature loudspeakers are placed on top. Tracking of the pen is realized in Processing through the Microsoft Kinect v2 sensor, covering an A2-sized sheet of paper, fixated on the table.

During the event, volunteers were asked to trace back the regions of different materiality through natural interaction either with fingers or a ball-pen, both producing sufficient auditory feedback to drive the auditory augmentation. The ball-pen additionally provides an intuitive way for switching between “exploration” and “drawing” mode through (de)activation of the tip.

Due to the lower resolution of the Kinect-based tracking compared to the graphic tablet, the fine-grained “Mondrian-style” structures were regarded as too difficult to detect in the envisaged context. Instead, users were asked to detect three different shapes, a rectangle, a triangle, and a circle, and to draw them on the paper sheet. The same materials as in the informal evaluation were mapped to the three shapes: rectangle to metal, triangle to ceramics, and circle to glass. The spatial arrangement of the shapes randomly differed between participants. The shapes could overlap, i.e., partially mask others. An example-image with the correct proportions was shown to the participants. Most of the time, the room was crowded with people, inducing a relatively noisy and sometimes disturbing environment.

20 random visitors were documented through their individual drawings. Four of these drawings are shown in Fig. 7, with the auditorily displayed arrangement drawn as an overlay. An analysis of

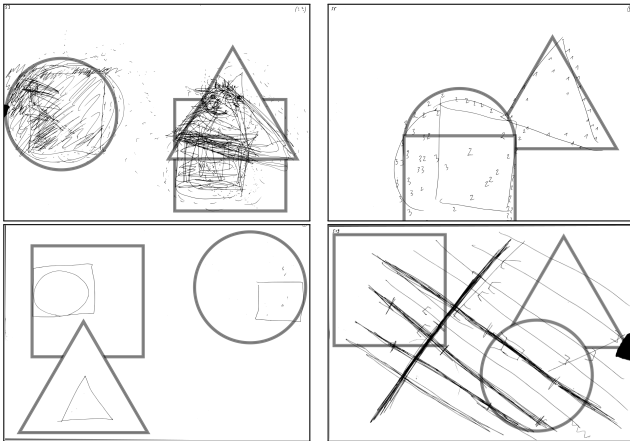


Figure 7: Drawings from participants 3 (top left), 5 (top right), 8 (bottom left), and 19 (bottom right), overlain by the corresponding “correct answer”, respectively.

these drawings through visual judgment of the investigator reveals that 55 % of the participants correctly identified the approximate location of all three regions. 40 % additionally assigned all correct shapes, while 60 % roughly hit the correct size. 95 % confidence intervals, assuming binomial distribution, are 32–77 %, 19–64 %, and 36–81 %, respectively.

Results indicate that the task was very hard to accomplish in the given context. However, the experiment clearly reveals 4 different strategies for task completion (compare Fig. 7): (1) continuous oscillatory drawing with enabled pen, somehow tracing back the border regions (top left), (2) tracing back the borders through systematic tapping, marking positions of different sound and connecting these (top right), (3) random tapping interaction to quickly find extreme values for efficient identification of the shapes (bottom left), and (4) continuous scanning of the surface and marking material changes (bottom right).

Yet interestingly, strategy (3) was only used by children and led to quick and reliable results for shape-identification; however, always underestimating the size. The other strategies concentrated on exact border positions, and took much longer (up to about 15 min), but gained similar hit-rate for shape-identification. The relatively low overall performance surely comes from low resolution and instability of the tracking system.

## 5. CONCLUSIONS AND OUTLOOK

We have introduced the research question of plausibility of auditory augmentations of physical interactions, with the particular interest in shedding light on the available bandwidth for sonification using auditory augmentation, understood as the number of discriminable alternative auditory feedbacks, so that the sound is still plausible and useful for the underlying physical interaction. This is basic research, and thus requires as a first step to clarify concepts, and to develop a research method to investigate the emerging questions.

With the developed experimentation platform, we have succeeded in creating settings that allow both a high-fidelity control of discrete and continuous physical interactions, and at the same

time to manipulate the auditory feedback from subtle to gross. This platform is intended to form the basis for conducting empirical studies in order to investigate limits of plausibility of auditory augmentations and how blended sonifications affect usability of the original auditory feedback.

The informal evaluation helped to develop the experimental platform as well as to design further experiments with it. The qualitative user study gave an insight on behavior and strategies of naive users of the system, which is of great importance for future designs. Especially time-varying auditory augmentations are supposed to work well for real-time sonification; however, presumingly at higher risk of inducing implausibility.

A series of follow-up studies will focus on the discovery of just-noticeable differences in auditory augmentation. A marker-based motion-capture system will be used in order to enable the auditory augmentation of arbitrary flat surfaces such as existing tables or plates of different material with high-precision tracking.

## 6. ACKNOWLEDGMENT

This research was supported by the Cluster of Excellence Cognitive Interaction Technology ‘CITEC’ (EXC 277) at Bielefeld University, funded by the German Research Foundation (DFG).

## 7. REFERENCES

- [1] N. Markosian, “What are physical objects?” *Philosophical and Phenomenological Research*, pp. 375–395, 2000.
- [2] V. Kaptelinin, “Context and consciousness: Activity theory and human-computer interaction,” B. A. Nardi, Ed. Cambridge, MA: MIT Press, 1995, ch. Activity Theory: Implications for Human-computer Interaction, pp. 53–59.
- [3] K. Kuutti, “Context and consciousness,” B. A. Nardi, Ed. Cambridge, MA, USA: Massachusetts Institute of Technology, 1995, ch. Activity Theory As a Potential Framework for Human-computer Interaction Research, pp. 17–44.
- [4] R. Bresin, A. de Witt, S. Papetti, M. Civolani, and F. Fontana, “Expressive sonification of footstep sounds,” in *ISon workshop*. KTH, 2010, pp. 51–54.
- [5] J. Maculewicz, C. Erkut, and S. Serafin, “An investigation on the influence of soundscapes and footstep sounds in affecting preferred walking pace,” in *ICAD*, 2015.
- [6] S. Papetti and F. Fontana, “Effects of audio-tactile floor augmentation on perception and action during walking: Preliminary results,” in *SMC*, 2012, pp. 17–22.
- [7] L. Turchet and R. Bresin, “Effects of interactive sonification on emotionally expressive walking styles,” *IEEE Trans. on Affective Computing*, vol. 6, no. 2, pp. 152–164, 2015.
- [8] E. Furfaro, F. Bevilacqua, A. Tajadura Jimenez, *et al.*, “Sonification of surface tapping: Influences on behaviour, emotion and surface perception.” *ISon Workshop*, 2013.
- [9] E. Furfaro, F. Bevilacqua, N. Berthouze, and A. Tajadura Jimenez, “Sonification of virtual and real surface tapping: evaluation of behavior changes, surface perception and emotional indices,” *IEEE MultiMedia*, 2015.
- [10] U. Castiello, B. L. Giordano, C. Begliomini, C. Ansuini, and M. Grassi, “When ears drive hands: the influence of contact sound on reaching to grasp,” *PLoS one*, vol. 5, no. 8, 2010.

- [11] A. Sedda, S. Monaco, G. Bottini, and M. A. Goodale, "Integration of visual and auditory information for hand actions: preliminary evidence for the contribution of natural sounds to grasping," *Experimental brain research*, vol. 209, no. 3, pp. 365–374, 2011.
- [12] T. Bovermann, R. Tünnermann, and T. Hermann, *Auditory Augmentation*, ser. Premier reference source, K. Curran, Ed. Information Science Reference, 2010.
- [13] F. Truffaut, A. Hitchcock, and H. G. Scott, *Hitchcock*, revised ed. Simon and Schuster, 1985.
- [14] L. Connell and M. T. Keane, "A model of plausibility," *Cognitive Science*, vol. 30, no. 1, pp. 95–120, 2006.
- [15] S. Serafin, H. P. Kjaer, C. Taylor, and F. Avanzini, "Audio-haptic physically based simulation and perception of contact textures," in *ICAD*, 2007.
- [16] R. Martín, J. Iseringhausen, M. Weinmann, and M. B. Hullin, "Multimodal perception of material properties," in *ACM SIGGRAPH on applied perception*, 2015, pp. 33–40.
- [17] B. L. Giordano and S. McAdams, "Material identification of real impact sounds: Effects of size variation in steel, glass, wood, and plexiglass plates," *JASA*, vol. 119, no. 2, pp. 1171–1181, 2006.
- [18] D. Rocchesso and L. Ottaviani, "Can one hear the volume of a shape?" in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 2001, pp. 115–118.
- [19] R. A. Lutfi, "Auditory detection of hollowness," *JASA*, vol. 110, no. 2, pp. 1010–1019, 2001.
- [20] R. Mason, T. Brookes, and F. Rumsey, "Frequency dependency of the relationship between perceived auditory source width and the interaural cross-correlation coefficient for time-invariant stimuli," *JASA*, vol. 117, no. 3, pp. 1337–1350, 2005.
- [21] D. R. Perrott and T. N. Buell, "Judgments of sound volume: Effects of signal duration, level, and interaural characteristics on the perceived extensity of broadband noise," *JASA*, vol. 72, no. 5, pp. 1413–1417, 1982.
- [22] D. Cabrera and S. Tilley, "Parameters for auditory display of height and size," in *ICAD*, 2003.
- [23] C. Marquis-Favre and J. Faure, "Auditory evaluation of sounds radiated from a vibrating plate with various viscoelastic boundary conditions," *Acta Acustica united with Acustica*, vol. 94, no. 3, pp. 419–432, 2008.
- [24] D. Rocchesso and F. Fontana, Eds., *The Sounding Object*. PHASAR, 2003. [Online]. Available: [http://www.soundobject.org/SObBook/SObBook\\_JUL03.pdf](http://www.soundobject.org/SObBook/SObBook_JUL03.pdf)
- [25] W. A. Yost, R. R. Fay, and A. N. Popper, *Auditory perception of sound sources*. Springer, 2007, vol. 29.
- [26] R. Wages, S. M. Grünvogel, and B. Grützmaker, "How realistic is realism? considerations on the aesthetics of computer games," in *International Conference on Entertainment Computing*. Springer, 2004, pp. 216–225.
- [27] ISO, "9241-210: 2010. ergonomics of human system interaction-part 210: Human-centred design for interactive systems," 2009.
- [28] R. Sigrist, G. Rauter, R. Riener, and P. Wolf, "Augmented visual, auditory, haptic, and multimodal feedback in motor learning: a review," *Psychonomic bulletin & review*, vol. 20, no. 1, pp. 21–53, 2013.
- [29] Z. Ma, D. Edge, L. Findlater, and H. Z. Tan, "Haptic keyclick feedback improves typing speed and reduces typing errors on a flat keyboard," in *World Haptics*, 6 2015, pp. 220–227.
- [30] D. Norman, *The design of everyday things*. Basic, 2002.
- [31] R. Tünnermann, J. Hammerschmidt, and T. Hermann, "Blended sonification – sonification for casual information interaction," in *ICAD*, 2013.
- [32] S. Barrass and T. Barrass, "Embedding sonifications in things," in *ICAD*, 2013.
- [33] S. Ferguson, "Sonifying every day: Activating everyday interactions for ambient sonification systems," in *ICAD*, 2013.
- [34] C. Müller-Tomfelde and T. Münch, "Modeling and sonifying pen strokes on surfaces," in *DAFx*, 2001.
- [35] T. Hermann, A. Hunt, and J. G. Neuhoff, *The sonification handbook*. Logos, 2011.
- [36] Y.-H. Yang and S.-L. Yeh, "Unmasking the dichoptic mask by sound: spatial congruency matters," *Experimental brain research*, vol. 232, no. 4, pp. 1109–1116, 2014.
- [37] J. Blauert, *Spatial Hearing: The Psychophysics of Human Sound Localization*, revised ed. MIT Press, 1997.
- [38] M. E. Altinsoy, "The quality of auditory-tactile virtual environments," *JAES*, vol. 60, no. 1/2, pp. 38–46, 2012.
- [39] B. D. Adelstein, D. R. Begault, M. R. Anderson, and E. M. Wenzel, "Sensitivity to haptic-audio asynchrony," in *ACM ICMI*, 2003, pp. 73–76.
- [40] D. J. Levitin, K. MacLean, M. Mathews, L. Chu, E. Jensen, and D. M. Dubois, "The perception of cross-modal simultaneity (or "the greenwich observatory problem" revisited)," in *AIP Conference*, vol. 517, no. 1, 2000, pp. 323–329.
- [41] A. McPherson and V. Zappi, "An environment for submillisecond-latency audio and sensor processing on beaglebone black," in *AES Convention*, 2015.
- [42] V. Zappi and A. McPherson, "Design and use of a hackable digital instrument," in *Proc. Live Interfaces*, 2014.
- [43] J. D. Tillman. (2001, 5) A discrete fet guitar preamp. [Online]. Available: <http://www.till.com/articles/GuitarPreamp/>
- [44] P. R. Cook, *Real sound synthesis for interactive applications*. CRC Press, 2002.
- [45] N. H. Fletcher and T. D. Rossing, *The Physics of Musical Instruments*, 2nd ed. Springer, 1998.
- [46] M. Aramaki, M. Besson, R. Kronland-Martinet, and S. Ystad, "Controlling the perceived material in an impact sound synthesizer," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 2, pp. 301–314, 2011.
- [47] G. Warburton, "The vibration of rectangular plates," *Proceedings of the Institution of Mechanical Engineers*, vol. 168, no. 1, pp. 371–384, 1954.
- [48] J. Fan. Mondrian generator. [Online]. Available: <https://github.com/JEFworks/mondrian-generator>
- [49] M. Weger. (2018, 2) Supplementary material. [Online]. Available: [https://phaidra.kug.ac.at/detail\\_object/o:69732](https://phaidra.kug.ac.at/detail_object/o:69732)